PB85-114072

Are the 'National Guidelines' Based on Sound Judgments

(U.S.) Environmental Research Lab.-Duluth, MN

Oct 84

U.S. Department of Commerce
National Technical Information Service
NTIS

PB85-114072

EPA-600/D-84-251
October 1984

Are the "National Guidelines" Based on Sound Judgments?

by

Charles E. Stephan

U.S. Environmental Protection Agency

Environmental Research Laboratory-Duluth

6201 Congdon Boulevard

Duluth, Minnesota 55804

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO.	2.	3. RECIPIENT'S ACCESSION NO.
EPA-600/D-84-251		PB85 114072

4. TITLE AND SUBTITLE	5. REPORT DATE
Are the "National Guidelines" Based on Sound Judgments?	October 1984
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
C. E. Stephan	
9. PERFORMING ORGANIZATION NAME AND ADDRESS	10. PROGRAM ELEMENT NO.
Environmental Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Duluth, MN 55804	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS	13. TYPE OF REPORT AND PERIOD COVERED
same as above	14. SPONSORING AGENCY CODE EPA-600/03

15. SUPPLEMENTARY NOTES

16. ABSTRACT

Until recently, procedures used to derive water quality criteria for aquatic life were not well defined and few principles were identified. On November 28, 1980, the United States Environmental Protection Agency published "Guidelines for Deriving Water Quality Criteria for the Protection of Aquatic Life and Its Uses" in the Federal Register. These have been subsequently revised and renamed to "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Life and Its Uses" and are referred to as the "National Guidelines." In addition, guidelines have been developed for deriving site-specific criteria either by modifying national criteria or by using other appropriate information. Establishing procedures for deriving water quality criteria and for assessing hazard to aquatic life have many similarities because both make use of information from many areas of aquatic toxicology and both assume that the science has developed sufficiently that these activities are feasible and desirable. The desirability of National Guidelines depends on the appropriateness of the strategy developed for using the resulting criteria and the numerous technical judgments that must be made when developing the Guidelines.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release to public	unclassified	26
	20. SECURITY CLASS *(This page)*	22. PRICE
	unclassified	

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT: Until recently, procedures used to derive water quality criteria for aquatic life were not well defined and few principles were identified. On November 28, 1980, the United States Environmental Protection Agency published "Guidelines for Deriving Water Quality Criteria for the Protection of Aquatic Life and Its Uses" in the Federal Register. These have been subsequently revised and renamed to "Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Life and Its Uses" and are referred to as the "National Guidelines." In addition, guidelines have been developed for deriving site-specific criteria either by modifying national criteria or by using other appropriate information. Establishing procedures for deriving water quality criteria and for assessing hazard to aquatic life have many similarities because both make use of information from many areas of aquatic toxicology and both assume that the science has developed sufficiently that these activities are feasible and desirable. The desirability of National Guidelines depends on the appropriateness of the strategy developed for using the resulting criteria and the numerous technical judgments that must be made when developing the Guidelines.

KEY WORDS: aquatic toxicology, water pollution, water quality criteria, acute-chronic ratio, bioconcentration, bioaccumulation

Most aquatic toxicologists are familiar with the colorful history of water quality criteria for aquatic life as exemplified in the Green Book [1], the Blue Book [2], and the Red Book [3]. Criteria in these books were derived by a variety of procedures, but the general approach might best be called the "lowest number approach" or "most sensitive species approach." Most of the criteria were based on the lowest available result from a toxicity test or were designed to protect the most sensitive species that had been tested. In January, 1978, when the U.S. EPA was preparing the sequel to the Red Book, Don Mount convinced appropriate people in the agency that there ought to be a better way to derive criteria. Naturally, the first step was to form a committee, and so six representatives from U.S. EPA's Environmental Research Laboratories began developing guidelines for deriving water quality criteria. One version was published in the Federal Register on May 18, 1978, [4] for public comment; another on March 15, 1979 [5]; and another on November 28, 1980, with response to public comment [6]. Since then, work has been progressing on a new version which will be titled "Guidelines for Deriving National Water Quality Criteria for the Protection of Aquatic Life and Its Uses" and will be available for public comment in 1983. The U.S. EPA has also proposed "Guidelines for Deriving Site-Specific Water Quality Criteria for the Protection of Aquatic Life and Its Uses" [7]. (Although these are commonly referred to as the National Guidelines and Site-Specific Guidelines, respectively, only the National Guidelines will be discussed herein, and they will be referred to simply as the Guidelines.)

The title of this article is not intended to mislead anyone into thinking that my answer to the question might be "No." Rather the title is intended to encourage people to realize that the Guidelines are based on numerous judgments, some of which are philosophical and some technical. My

purpose here is to promote consideration of some of the judgments underlying the Guidelines.

In many respects, developing guidelines for deriving water quality criteria is very similar to writing a standard practice for assessing hazard to aquatic organisms. One of the obvious similarities is that some of us have been working on both of these for what seems to be a long time. The major similarity, however, is that both require consideration of many facets of aquatic toxicology and both require numerous judgments concerning generalities as well as specifics. If the major difference between hazard assessment and risk assessment is that hazard assessment is qualitative and risk assessment is quantitative, then deriving numerical water quality criteria can be considered a form of risk assessment rather than hazard assessment. It is illuminating to consider the similarities between deriving water quality criteria and assessing hazard because many of the same philosophical and technical decisions have to be made in both activities. Although much effort has been spent in the last few years in ASTM on a practice for assessing hazard and in U.S. EPA on guidelines for deriving criteria, nobody expects the final word to come soon in either area because both involve working on state-of-the-art issues in aquatic toxicology. Work in both areas continues because people feel that both are feasible and desirable, even though the questions of feasibility and desirability have not been examined very closely.

Feasibility of National Guidelines

Criteria presented in the Green Book, Blue Book, and Red Book were derived using whatever data were available and whatever rationale was considered appropriate for interpreting the data that were available for each

individual material. A basic judgment underlying the Guidelines is that a valid, comprehensive procedure can be applied to all materials. Note carefully that the claim is that the Guidelines can be applied to all materials; the claim is not that the Guidelines will allow derivation of criteria for all materials. Reasonable Guidelines must acknowledge at a number of points that for some materials the available data may not fit a recognizable pattern and so it may not be possible to derive a water quality criterion for aquatic life. In spite of differences between materials, however, it should be possible to develop a comprehensive procedure that will be valid for all materials.

Another basic judgment is that not only is the available information sufficient to allow us to envision Guidelines, enough information is presently available to develop the Guidelines. Even though all of the desirable information is not available, the data that are available provide an adequate basis for the Guidelines. When deriving water quality criteria or assessing hazard, decisions are based on data as much as possible, but in almost every case it is necessary to choose between conflicting data, to adopt simplifications, or to go beyond the available data. New data usually both answer and raise questions. Even if limited resources were not a problem, there will always be unanswered questions. Aquatic toxicologists will always be faced with the desire for more data. Thus a fundamental judgment underlying the Guidelines is that, in spite of a variety of unanswered questions, aquatic toxicology has advanced to the point that adequate information exists in the pertinent areas to develop guidelines for deriving water quality criteria for aquatic life.

A corollary of this judgment about the state-of-the-art of aquatic toxicology is that the Guidelines are not "cast in stone". Much desired

information is not available; therefore as new information and better rationales are developed, changes will be necessary. A major side benefit of the effort to develop Guidelines is that it aids in the development of new data by causing the re-examination of available data, the proposal of new ideas, and the clarification of research needs. Although new data and ideas should result in improvements from time to time, current information certainly justifies the development of Guidelines at this time.

Thus the Guidelines are predicated on two fundamental judgments: one concerning the basic applicability of general principles to most materials and the other concerning the state-of-the-art of aquatic toxicology that lead to the conclusion that Guidelines are feasible. An additional but equally important question is whether they are desirable.

Desirability of Guidelines

Two of the more fundamental problems in aquatic toxicology are that (a) water quality can affect the toxicity of most materials and (b) aquatic species show a range of sensitivities to most materials [8]. It would seem only logical, therefore, that national criteria are useless because the only good criteria are site-specific criteria. Although the authors of the Guidelines realize the importance of local or site-specific criteria, the rationale of the relationship of national criteria to site-specific criteria has developed from a vague concept in 1978 to a more well-defined idea in 1980 to a specific strategy in 1983. The assumption is that if national criteria are appropriately derived, both the need for and the cost of deriving site-specific criteria can be minimized. The strategy is intended to be cost-effective, i.e., to minimize costs associated with site-specific criteria, by ensuring that most, but not necessarily all, site-specific

criteria for a material are higher than the national criterion for the material.

This is a cost-effective strategy because it permits the assumption that if the concentration of a material in a body of water is lower than the national criterion, the aquatic life usually will not be unacceptably affected; thus neither a site-specific criterion nor additional pollution control is needed. This means that site-specific criteria do not have to be derived for most materials in most bodies of water in which the actual concentrations do not exceed national criteria. Any other approach to the relationship of national criteria to site-specific criteria would mean that a site-specific criterion would have to be derived for each body of water in which there was any concern about the concentration of a particular material.

However, in order for this strategy to be cost-effective, the Guidelines must not only result in national criteria that are not too high, they must also result in criteria that are not too low. If national criteria are unnecessarily low, too many site-specific criteria will have to be derived. In an attempt to be low enough but not too low, a national criterion is intended to be an appropriate criterion for an aquataic community that is among the most sensitive to the material of concern in water that contains low concentrations of substances that can reduce the toxicity of the material. If the highest acceptable concentration of a material were known for all bodies of water in the United States, the national criterion for the material would be equal to the lowest of these concentrations that was not judged to be an outlier.

The second part of the strategy is that by using appropriate procedures for determining the relative sensitivities of various aquatic communities and

the relative toxicities of a material in different waters, it is possible to derive many site-specific criteria merely by modifying national criteria. The strategy, therefore, not only reduces the number of site-specific criteria that are needed, but it also reduces the cost of obtaining many of the site-specific criteria that are needed.

Another way in which valid Guidelines will save money is by resulting in better national criteria and better site-specific criteria. If either kind of criteria are derived by different people using different procedures, at least some of the criteria will be much too high and some will be much too low. If criteria are too low, money will be unnecessarily spent on pollution control and the nation will suffer economically. On the other hand, aquatic life will suffer if criteria are too high; and if aquatic life suffer too much, the nation will also suffer. Therefore, it is in the best interest of the nation that criteria be neither excessively high nor excessively low.

The important point is that the community of aquatic toxicologists must decide whether national water quality criteria for aquatic life should be derived using some form of Guidelines or whether they should be derived without Guidelines. The Guidelines are based on the dual judgments that Guidelines are both feasible and desirable because valid Guidelines will be in the best interest of both the nation and the aquatic life. Although the questions of feasibility and desirability are partly technical and partly social, aquatic toxicologists must consider them to keep their work in perspective. Developing guidelines for deriving criteria and developing practices for assessing hazard must be both technically feasible and socially desirable if they are to be accepted as useful activities.

If the strategy for national and site-specific criteria is to be cost-effective, it is not enough to derive the best possible criterion for a

material; it is equally important that each criterion be a good estimate. If the concept of the Guidelines were merely to derive the best possible criterion based on available data, the Guidelines would provide ways of interpreting whatever data were available. In the extreme, if no results of toxicity tests or bioconcentration tests were available for a material, criteria for that material would be derived by extrapolation based on data on physical and chemical properties, structure, data on related materials, or some combination of all three. Criteria derived merely by doing the best that can be done using existing data are likely to sometimes be much too high and sometimes be much too low, and such criteria would not be cost-effective.

To help ensure that criteria are generally good estimates, the concept of required data has been incorporated into the Guidelines. The idea was to define the required data so that when all are available, a good criterion can usually be derived; whereas when all the required data are not available, a criterion usually should not be derived. The distinction between the qualitative process of assessing hazard and the quantitative process of deriving criteria is pertinent here because more data are necessary to be quantitative than to be qualitative. Data currently required by the Guidelines are:

1. acute tests with species in at least eight different families;
2. acute-chronic ratios with at least three species;
3. a test with at least one plant species; and
4. a bioconcentration factor in some cases.

The concept of required data is a means of implementing the idea that criteria can only be cost-effective if they are good estimates. This mechanism is obviously not an ideal solution because good criteria cannot be

derived from some sets of data which contain all the required data and, alternatively, good criteria can be derived from some sets that do not contain all the required data. It would be desirable to have a better means of implementing the concept of "good criteria," but as yet the state-of-the-art has not advanced that far.

What do the Guidelines Intend to Protect?

It would seem that one of the first steps in assessing hazard or deriving water quality criteria for protecting aquatic life would be to develop a reasonably clear definition of what is meant by "protection". Unfortunately, neither the ASTM drafts on assessing hazard nor the published versions of the Guidelines have dealt with this issue. Over the years many people have expressed opinions about what constitutes adequate protection and the ideas cover a wide gamut. For this reason most aquatic toxicologists probably feel it is prudent to avoid the topic as long as possible. The proper attitude, of course, is that a cost-effective strategy for achieving protection of aquatic life is the ultimate goal of aquatic toxicology and therefore toxicologists must conscientiously work at defining the concept of protection if there is to be any justification for everything else that they do. In order to convince other scientists and the public that aquatic toxicology is a useful activity, aquatic toxicologists must work seriously at defining what constitutes adequate protection.

In the water quality-based approach to pollution control, the public decides what uses are to be protected in each body of water. If the use known as "aquatic life" is to be protected, then criteria necessary to protect that use must be incorporated into the water quality standards. One of the important aspects of protection of aquatic life is that it is not enough to protect the presence of aquatic life; its uses must also be

protected. The uses are important. If water quality criteria for aquatic life are not designed to protect the uses of aquatic life, the uses may not be protected. For example, many commercially and recreationally important aquatic species will be useless if they taste so bad that nobody will eat them or if they contain concentrations of materials that exceed FDA action levels. Recreational and commercial fisherman will be quite unhappy if they cannot eat or sell their catch after aquatic toxicologists have said that aquatic life is adequately protected.

Another important aspect of protection is that there are very few kinds of aquatic species about which the public is concerned. Judging by the things that people usually complain about, as long as the presence and uses of these few species are not noticeably affected, most people feel that aquatic life are adequately protected. There are both fish and invertebrate species in salt water about which the public is concerned, but most people only care about fish in fresh water. Even so-called extremists - the snail darter types - are named after a fish.

Very few people are concerned about adverse effects on various species of aquatic bacteria, fungi, protozoans, phytoplankton and zooplankton unless effects on those species result in unacceptable effects on a commerically or recreationally important species. Thus, real world protection of aquatic life and its uses should be operationally defined in terms of field monitoring of the kinds of species that the public cares about. Such monitoring would of course have to continue for many years to take into account seasonal and annual fluctuations. Also, it would be impossible to adequately monitor some species and would be prohibitively expensive for many others. The practical alternative would be to monitor surrogate species. Appropriate monitoring of desirable and surrogate kinds of species would

detect any direct unacceptable effects and would also detect any indirect unacceptable effects that might be caused by such things as loss of a key food organism, a change in energy flow, or a change in a predator-prey relationship. Such monitoring should be designed to detect effects that the public would consider unacceptable, rather than trying to detect other kinds of effects and extrapolating to effects that the public would consider unacceptable. If appropriately performed on a regular and continuing basis, a well designed monitoring program would detect unacceptable effects regardless of whether they were caused directly or indirectly. My purpose here is not to discuss monitoring programs per se but to express the point of view that an appropriate definition of "protection of aquatic life and its uses" should be based on the kinds of species that most people actually care about.

Although this kind of definition of "protection of aquatic life and its uses" will certainly be considered unacceptable by some people, in my opinion the primary goal of aquatic toxicology should not be to protect such things as the function or structure of aquatic ecosystems unless effects on such things result in unacceptable effects on species of concern to the public. In particular, there is no reason to protect a species of bacteria, fungi, phytoplankton or zooplankton if that species can be replaced by one or more other species so that the kinds of species that most people care about are not adversely affected. This concept of protection leads directly to two judgments that are used in the Guidelines.

An unstated concept is that the loss of a few or even several species among the lower forms of life, such as bacteria, fungi, and protozoans, is not of concern because there are so many other species that can replace the ones that are lost. In most cases a human-induced shift in a species

composition of some lower forms of life will not cause a problem. Also, most lower forms reproduce so rapidly that adaptation and repopulation can take place quickly. This may apply even to species as high as algae. For example, in the Shayler Run study [9], the test concentration of copper decimated the dominant algal species. However, other species replaced it so well that algal biomass was not reduced, algal diversity increased [10] and no resulting adverse effects on the fish and macroinvertebrates were detected. Eliminating all species of bacteria, fungi, protozoans or algae in a body of water would probably cause unacceptable effects on important species, but harming only a few such species, apparently including some dominant species, will not always cause a problem.

A second fundamental judgment, which is stated in the Guidelines, concerns the number of higher species that need to be protected. The rationale is that it is not necessary to protect all species all the time. Aquatic communities can recover from some short-term adverse effects and they can adapt to some long-term adverse effects. The approach generally used in the Green Book, Blue Book, and Red Book was to set criteria so that all species that had been tested would be protected. This approach is usually criticized as resulting in criteria that are too low, but the resulting criteria can be too high if the most sensitive tested species is not as sensitive as some important species. In general, however, this approach will usually be overprotective. The major advantage of this approach is that it is fairly easy to search the literature, find the lowest number, and use it as the criterion. Unfortunately, although it is easy to decide that it is not always necessary to protect all species, this decision raises many difficult questions. The judgment used in the Guidelines is that if acceptable data are available on the toxicity of a material to a variety of

appropriate species, the criterion should be set to protect (a) 95 percent of the tested species and (b) all commercially, recreationally, and socially important species.

Qualitative and Quantitative Judgments

As a brief but important digression, let me comment on the vast difference between making philosophical or qualitative judgments and making quantitative judgments. Whereas philosophical and qualitative judgments can be based on rationales, many scientists seem to feel that quantitative judgments must be based on data. Obviously, however, if the appropriate data were available, the desired number could be calculated and the issue would not have to be decided by judgment. The counter argument is that if the necessary data are not available, judgment should not be used as a substitute for data. This is as unrealistic in applied aquatic toxicology as it is in most areas of life. Making assumptions and simplifications is a fact of life. The two times when judgment is clearly inappropriate is when it contradicts the available data and when the alternatives cover such a broad range that any decision is merely a guess and may be very unrealistic. Under reasonable circumstances, however, quantitative judgments can be just as justified as qualitative judgments.

The major problem with quantitative judgments, as opposed to qualitative or philosophical judgments, is that people always ask "Why 95? Why not 94 or 96? or "Why 8? Why not 7 or 9?" The problem, of course, is that because it is a judgment, adequate data are not available to quantitatively justify the decision. For example, 95 was chosen because 90 and 99 resulted in Final Acute Values that seemed to be too high and too low, respectively, when compared to the data sets from which they were calculated. Of the numbers

available between 90 and 99, 95 is near the middle and is an easily recognizable number. On the other hand, 8 was chosen because 8 acceptable values were available for many materials, but more than 8 were available for only a few. Although there is not much of a difference between 7, 8, and 9, all of these are quite different from numbers like 1, 2, and 3, which have been advocated by some people. Hopefully, advances in aquatic toxicology will allow better justifications for these or other numbers, but at this time the Guidelines describe the best available way of deriving national criteria; in addition, it is felt that these criteria are a useful basis for a cost-effective strategy for protection of aquatic life.

All of the problems of quantitative judgments could be avoided by not putting that level of detail into the Guidelines. This would give users of the Guidelines lots of flexibility to make appropriate case-by-case decisions. Unfortunately, even experienced aquatic toxicologists have different viewpoints on major and minor decisions. The fewer details there are in the Guidelines, the more variation there will be between criteria derived by different people for the same material. Thus it was necessary to make the Guidelines as detailed as feasible. The best way to decide what level of detail is appropriate in the Guidelines is to listen to the questions asked by people who try to use the Guidelines to derive criteria for aquatic life. The conclusion is that it is difficult to include too much detail. People who favor more flexibility and less detail usually feel either (a) that the Guidelines are biased toward underprotection or overprotection (both claims have been made) or (b) that some situations are so abnormal that predetermined Guidelines cannot adequately deal with them. The latter argument is the reason that some of the details in the Guidelines are explanations as to why criteria should not be derived in some specific

situations. The Guidelines do specify that the most important tenant is that criteria should be based on good science. When good science and the Guidelines do not agree, good science must be followed. Good science is certainly a valid reason for not following the Guidelines, but individual whim is not. There is a big difference between a position based on good science and one based on a personal preference.

Two-number Criteria

One of the major new features of the Guidelines is that a water quality criterion for aquatic life should consist of more than one number. The Blue Book mentioned the idea of two-number criteria, but never actually derived any such criteria. Later, John Eaton [11] proposed values for two-number criteria for some pesticides but the proposal was never adopted. Organisms can usually tolerate higher concentrations for short periods of time than they can for long periods and very few discharges are constant quality. If a never-to-be-exceeded, one-number criterion adequately protects aquatic life from long-term exposures, it will over-restrict dischargers by prohibiting short-term higher concentrations that could be tolerated by aquatic life. Similarly, a one-number average criterion will either underprotect aquatic life or overly restrict dischargers. In the worst of all possible cases, a criterion would both overly restrict dischargers and not adequately protect aquatic life.

The easy decision to have more than one number in criteria results in the very difficult problem of how to do it. All kinds of combinations of two or more numbers and time periods can be proposed, including graphs. To be realistic, however, an approach must take into account (a) the kinds of toxicological data that are, or are likely to be, available; (b) the

differences between aquatic species; and (c) the practicalities of treatment plant operation and monitoring programs faced by dischargers and regulatory agencies. The simplest alternative to one-number criteria is, of course, two-number criteria and so the Guidelines specify criteria in terms of an average concentration and a maximum concentration. This is judged to be the best that can be done with the kinds of data that are, or are likely to be, available; in addition, a two-number criterion can adequately protect aquatic life without being unfair to dischargers. There are many formats that could be used for two-number criteria and many ways the two numbers might be calculated, so adoption of the idea of a two-number criterion still presents many options. The option used in the Guidelines is intended to be the best way of using available data to obtain the best two numbers that will provide reasonable flexibility to dischargers while also protecting aquatic life from exposures to long-term average concentrations and short-term exposures to higher concentrations.

The question of the number of numbers in a criterion is tied directly to the issue of how criteria are used. Criteria do not limit dischargers. Dischargers are limited by effluent limitations, which are sometimes calculated from water quality standards, which in turn are based on water quality criteria. Extrapolating from a criterion to an effluent limitation can be technically complicated, and legal, economic, and social considerations often magnify the level of difficulty. Even if criteria are derived appropriately, standards may be inappropriate if, for example, the wrong use is selected to be protected. Further, even if the standard is appropriate, the effluent limitation may provide more or less protection than needed by the aquatic life. Two-number criteria are derived on the assumption that a discharger might want to use flow-proportional discharge in

order to discharge the maximum concentration allowed by the criterion during each period of time. Most dischargers and regulatory agencies do not consider flow-proportional discharge a viable option, and so permit limitations usually allow the discharge of some amount less than the theoretical maximum. The processes of deriving standards and effluent limitations are possibly as complex as the process of deriving water quality criteria; they have to deal with economic and social, as well as technical issues, and these can have as much bearing on the actual amount of protection afforded aquatic life as the Guidelines. Many of the important judgments concerning how much protection is afforded aquatic life are outside the realm of the Guidelines.

Acute-Chronic Ratios

One of the enduring controversies in aquatic toxicology is the appropriateness of using application factors. The Guidelines cleverly avoid this issue by using acute-chronic ratios instead of application factors. The problem is still the same, however, and the Guidelines permit use of an acute-chronic ratio to derive criteria for a particular material only if enough data are available for that material to justify its use. Because aquatic toxicology cannot yet provide a general answer to this problem, the Guidelines wisely require that at least a minimum amount of pertinent data be available and that the decision be based on data, not theories. Although the original suggestion was that experimentally determined application factors for a material would be the same for all species of fish [12], it has been found that acute-chronic ratios experimentally obtained for a material using different species of fish and invertebrates often increase or decrease as the acute sensitivities of the species increase or decrease. Because criteria

are based on the idea of protecting 95% of the tested species, the acute-chronic ratio used must be one that is appropriate to the fifth percentile. The Guidelines place quite stringent limitations on the use of acute-chronic ratios.

The subject of acute-chronic ratios raises a minor but interesting point. Some of the public comment in 1978 stated that geometric means were used instead of arithmetic means in various places in the Guidelines merely to get a lower number. Although it is true that the geometric mean of a set of numbers will always be lower than the arithmetic mean, it is not true that use of the geometric mean will always result in a lower criterion. In addition, there is usually at least one mathematical rationale for choosing between a geometric mean and an arithemtic mean. As an illustration, assume that both acute and chronic tests have been conducted on a material with four different species with the following results:

Species	Acute Value (µg/litre)	Chronic Value (µg/litre)	Acute-Chronic Ratio	Application Factor
A	0.6400	0.0800	8.000	0.1250
B	1000	100.0	10.00	0.1000
C	320.0	20.00	16.00	0.0625
D	20.00	1.000	20.00	0.0500
Arithmetic Mean			13.50	0.0844
Geometric Mean			12.65	0.0791

If the acute value for another species is 100 µg/litre, what is the best estimate of its chronic value? Four calculations are possible using the four means:

	Acute-Chronic Ratio	Application Factor
Arithmetic Mean	$\frac{100\ \mu g/litre}{13.50} = 7.41\ \mu g/litre$	100 µg/litre x 0.0844 = 8.44 µg/litre
Geometric Mean	$\frac{100\ \mu g/litre}{12.65} = 7.91\ \mu g/litre$	100 µg/litre x 0.0791 = 7.91 µg/litre

Note that the answer is the same using the two geometric means, but not the two arithmetic means. This is one reason why it is usually best to use geometric means rather than arithmetic means when dealing with ratios and similar kinds of data. On the other hand, the statistical reason is that ratios are more likely to be lognormally distributed than normally distributed. In this example the way to get the lowest possible criterion would be to use arithmetic means with acute-chronic ratios, except that if application factors are used, then geometric means would give the lowest criterion. Decisions concerning the content of the Guidelines should not be based on an attempt to make the resulting criteria as low or as high as possible.

Final Residue Value

Two of the most obvious ways in which the Guidelines might result in Final Residue Values that are too high could be avoided if the list of required data were strengthened. The judgment was, however, that these shortcomings are not serious enough in most cases to make the criteria undesirable or to require additional expensive data. The first area of concern is that data on chronic effects are not available for many important species of wildlife consumers of aquatic life. Without such data, it is impossible to know whether various wildlife species might be unacceptably

affected by materials accumulated by aquatic life. The second area of concern is that bioaccumulation factors (BAFs) might be higher than bioconcentration factors (BCFs) for many materials. BCFs are determined in laboratory bioconcentration tests and are intended to measure only net uptake directly from water, although some additional uptake may occur if the food sorbs some of the material before it is eaten by the test organisms. The term BAF is used here to refer to the situation in which the food eaten by the organism is in steady-state with the concentration in the water so that the organisms of concern proportionately accumulate material from both food and water. Whereas a BCF almost has to be measured in a laboratory test, the best way to measure a BAF is in a field situation. For several materials BAFs appear to be higher than BCFs [13-18].

For many materials adequate data are not available concerning either toxicity to wildlife or BAFs or both. Thus for many materials the Final Residue Value either is too high or does not exist at all. It was decided, however, that if the required data were available, it would be better to derive criteria using the available data even if the Final Residue Value might be too high. Even some of the data that are available are not easy to use. Some wildlife studies report that the lowest concentration tested caused an adverse effect. Similarly, FDA action levels might be considered unacceptable concentrations because a Final Residue Value calculated from a BCF or BAF and an FDA action level should result in 50 percent of the organisms exceeding the FDA action level. More importantly, if the BCF or BAF is an average of values for different species, all the individuals of some species may exceed the FDA action level.

A common way of dealing with situations in which data are lacking or incomplete is to use safety or uncertainty factors. Mammalian toxicologists

routinely use factors of 10, 100, and 1000 [6], but such factors have not become accepted in aquatic toxicology. Safety factors are not used in the Guidelines because the implications of national criteria are so great that safety factors are not considered cost-effective and are not technically justifiable. When available data do not allow adequate confidence in a criterion, the only acceptable alternatives are either to obtain additional information or to not derive a criterion.

Summary

Because numerous judgments were made during the development of the Guidelines, this discussion has only dealt with the major philosophical issues that determine the overall nature of the Guidelines and with a few representative important technical issues to show how various kinds of decisions were made. In addition, the validity of a water quality criterion for a material depends just as much on the validity of numerous detailed technical decisions concerning that material as it does on the validity of the Guidelines. Anyone who tries to use the Guidelines quickly finds that criteria cannot be derived mechanically. Numerous "small" decisions must be made and some of these can substantially affect the resulting criterion. The Guidelines provide a framework for deriving criteria and they attempt to establish an attitude toward derivation of water quality criteria for aquatic life, but criteria still must be derived by people who are both conscientious and competent. It is to be hoped that a better understanding of the Guidelines will result in increased confidence in the resulting criteria and will help interested persons ask questions and make suggestions that will help improve the Guidelines and the resulting criteria. In addition to resulting in better national and site-specific criteria at less cost, the

Guidelines have resulted in a better understanding of the relationships between various areas of aquatic toxicology and have resulted in the formulation of ideas that ought to be tested.

Acknowledgments

The committee consisting of Don Mount, Dave Hansen, Jack Gentile, Gary Chapman, Bill Brungs and Charles Stephan developed the Guidelines, with input from many other people. Various people at the U.S. EPA's Environmental Research Laboratories in Corvallis, Oregon; Duluth, Minnesota; Gulf Breeze, Florida; and Narragansett, Rhode Island provided most of the input on the technical content of various aquatic life criteria documents. All of these people contributed to this paper, but none of them necessarily agree with anything contained herein.

References

[1] National Technical Advisory Committee, Water Quality Criteria, Federal Water Pollution Control Administration, Washington, D. C., 1968.

[2] National Academy of Sciences-National Academy of Engineering, Water Quality Criteria 1972, EPA-R3-73-033, U.S. Environmental Protection Agency, Washington, D. C., 1973.

[3] U.S. Environmental Protection Agency, Quality Criteria for Water, Washington, D. C., 1976.

[4] U.S. Environmental Protection Agency, Federal Register, Vol. 43, No. 97, May 18, 1978, pp. 21506-21518.

[5] U.S. Environmental Protection Agency, Federal Register, Vol, 44, No. 52, March 15, 1979, pp. 15926-15981.

[6] U.S. Environmental Protection Agency, Federal Register, Vol. 45, No. 231, November 28, 1980, pp. 79318-79379.

[7] U.S. Environmental Protection Agency, Federal Register, Vol. 47, No. 210, October 29, 1982, pp. 49234-49252.

[8] Stephan, C. E. in Aquatic Toxicology and Hazard Assessment, ASTM STP 766, American Society for Testing and Materials, Philadelphia, 1982, pp. 69-81.

[9] Weber, C. I. and McFarland, B. H. in Ecological Assessments of Effluent Impacts on Communities of Indigenous Aquatic Organisms, ASTM STP 730, American Socity for Testing and Materials, Philadelphia, 1981, pp. 101-131.

[10] Geckler, J. R., et al., "Validity of Laboratory Tests for Predicting Copper Toxicity in Streams," EPA-600/3-76-116, National Technical Information Service, Springfield, Va, December 1976.

[11] Eaton, J. G., Personal communication, U.S. EPA, Duluth, Minnesota.

[12] Mount, D. I. and Stephan. C. E., Transactions of the American Fisheries Society, Vol. 96, No. 2, April 1967, pp. 185-193.

[13] Macek, K. J., et al., Aquatic Toxicology, ASTM STP 667, American Society for Testing and Materials, Philadelphia, 1979, pp. 251-268.

[14] Bahner, L. H., et al., Chesapeake Science, Vol. 18, 1977, pp. 299-308.

[15] Jàrvinen, A. W., and Tyo, R. M., Archives of Environmental Contamination and Toxicology, Vol. 7, 1978, pp. 409-421.

[16] U.S. Environmental Protection Agency, "Ambient Water Quality Criteria for Polychlorinated Biphenyls," EPA-440/5-80-068, National Technical Information Service, Springfield, Va, 1980, pp. B7-B10.

[17] Boudou, A., et al., Bulletin of Enviornmental Contamination and Toxicology, Vol. 22, 1979, pp. 813-818.

[18] Phillips, G. R., and Buhler, D. R., Transactions of the American Fisheries Society, Vol. 107, 1978, pp. 853-861.